


Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ "НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМ. Н.И. ЛОБАЧЕВСКОГО"
(ННГУ)

УДК 004.8

№ госрегистрации АААА – А20 – 120111990045 – 1

УТВЕРЖДАЮ
Ректор ННГУ
д.м.н., профессор РАН, член –
корреспондент РАН




Е. В. Загайнова
«29» декабря 2022 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

по теме:

НАДЕЖНЫЙ И ЛОГИЧЕСКИ ПРОЗРАЧНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ:
ТЕХНОЛОГИЯ, ВЕРИФИКАЦИЯ И ПРИМЕНЕНИЕ ПРИ СОЦИАЛЬНО-ЗНАЧИМЫХ И
ИНФЕКЦИОННЫХ ЗАБОЛЕВАНИЯХ
(итоговый)

Подпрограмма «Фундаментальные научные исследования для долгосрочного
развития и обеспечения конкурентоспособности общества и государства»
государственной программы Российской Федерации
«Научно-технологическое развитие Российской Федерации»

Руководитель НИР,
Профессор кафедры нейротехнологий
Института биологии и биомедицины
д.ф.-м.н.

А.Н. Горбань

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	33
1. Доработка разработанных технологий и проведение их валидации на разнородных реальных данных, подготовленных при решении биомедицинских и других стратегически важных задач реальной сложности. Подготовка пользовательского программного обеспечения.	51
1.1. Изучение возможных скрытых (стелс) атак на системы ИИ и методов защиты от них	51
1.1.1. Введение.....	51
1.1.2. Скрытые атаки	53
1.1.3. Новые алгоритмы скрытой атаки.....	58
1.1.4. Эксперименты.....	69
1.1.5. Доказательства теорем и дополнительные результаты	82
1.1.6. Заключение.....	96
1.2. Обучение по небольшому числу примеров с нелинейно заданными свойствами.....	98
1.2.1. Введение.....	98
1.2.2. Проблема обучения с несколькими предъявлениями	101
1.2.3. Абстрактная теория подходов к обучению с несколькими предъявлениями	103
1.2.4. Геометрическая интерпретация результатов	107
1.2.5. Тестирование на обучении нейронных сетей	111
1.2.6. Заключение.....	114
1.3. Подготовка пользовательского программного обеспечения.....	114
1.3.1. Способ и система восстановления спектрального сверхразрешения ночных снимков городских агломераций	116
1.3.2. Способ распознавания изображений с обратимой коррекцией ошибок и интеллектуальная система для его реализации	124
1.3.3. Заключение.....	132
1.4. Первая теорема Розенблатта и бережливость глубокого обучения.....	133

1.4.1. Постановка задачи.....	133
1.4.2. Решение задачи с помощью полносвязной нейросети	134
1.4.3. Решение задачи с помощью глубокой нейросети	136
1.4.4. Нейросеть для -ограниченной задачи.....	136
1.4.5. Выводы и перспективы	137
1.5. Проведение суперкомпьютерной имплементации и тестирование новых алгоритмов оптимизации	137
1.5.1. Введение.....	137
1.5.2. Выбор оптимальных значений параметров модели склонового потока .	139
1.5.3. Использование методов редукции размерности для быстрой реализации алгоритмов машинного обучения	143
1.5.4. Заключение.....	145
1.6. Модель памяти интеллектуального агента, основанная на достижении нетривиальных предсказаний результатов действий	146
1.6.1. Введение.....	146
1.6.2. Алгоритм активного распознавания	147
1.6.3. Результаты и примеры работы	148
1.6.4. Нетривиальность предсказаний	151
1.6.5. Заключение.....	152
2. Проведение адаптации методов объяснимого и логически прозрачного ИИ в разработанные алгоритмы и программы для изучения механизмов болезней, ранней диагностики и идентификации возможных молекулярных и генетических мишеней для возможного лечения.	154
2.1. Введение	154
2.2. Данные метилирования ДНК.....	156
2.3. Мета-анализ и гармонизация.....	157
2.3.1. Методы	157
2.3.2. Результаты.....	159
2.4. Классификационные модели	160
2.4.1. Методы	160

2.4.2. Результаты.....	162
2.5. Импутация пропущенных значений	164
2.5.1. Методы	164
2.5.2. Результаты.....	166
2.6. Интерпретация и объяснимость	167
2.7. Заключение.....	173
3. Проведение имплементации методов коррекции ошибок и получение надежных систем ИИ с объясняемыми решениями для задач реальной сложности из различных областей науки. Тестирование системы на новых данных. Анализ результатов. Получение конкретных верифицируемых результатов в заявленные практические приложения.	
3.1. Имплементация методов коррекции ошибок и получение надежных системы ИИ с объясняемыми решениями; тестирование систем на новых данных; получено решение конкретных научных задач и получение прорывных, обусловленных развитыми методами результатов по областям	175
3.1.1. Анализ разнородных данных геномики, транскриптомики, метаболомики и биоинформатических сетей сложной структуры с целью идентификации молекулярных механизмов различных социально значимых заболеваний, разработки методов ранней диагностики и идентификации возможных молекулярных и генетических мишеней для возможного лечения.....	175
3.1.2. Моделирование сложных квантовых систем.....	202
3.1.3. Нахождение материалов, обладающих заранее заданными свойствами, методами хемоинформатики	212
3.1.4. Моделирование экспериментов в области сверхинтенсивных лазерных импульсов.....	216
3.1.5. Обработка многомерных и мультисенсорных сигналов и изображений для опережающей диагностики стрессовых состояний сельскохозяйственных растений	231
3.1.6. Разработка многомасштабных моделей нейронных сетей, реализующих информационно-вычислительные (интеллектуальные) функции мозга	242
3.1.7. Анализ климатических данных сложной структуры для оценки устойчивости климатической системы и для предсказания в ней критических событий	252
3.2. Будут разработаны экономные вычислительные алгоритмы для приближенного решения NP-задач по вычислению статистик на сетевых графах с	

помощью анализа распределения степеней вершин и величины следов степеней матрицы инцидентности. Будут разработаны методы приближенной оценки числа циклов сильно связной компоненты онтологических графов, с которыми связаны задачи повышения точности автоматического семантического анализа.....	265
3.2.1. Введение.....	265
3.2.2. Разработка экономных вычислительных алгоритмов для приближенного решения NP-задач по вычислению статистик на сетевых графах с помощью анализа распределения степеней вершин и величины следов степеней матрицы инцидентности.	267
3.2.3. Разработка методов приближенной оценки числа циклов сильно связной компоненты онтологических графов, с которыми связаны задачи повышения точности автоматического семантического анализа	310
3.2.4. Заключение.....	328
3.3. Развитие и адаптация приближенных, быстрых алгоритмов моделирования течений, позволяющих снизить сложность моделирования на порядок на базе метода POD/PCA на примере задачи с обтеканием цилиндра (профиля)	329
3.3.1. Введение.....	330
3.3.2. Обзор метода вихревых частиц.....	333
3.3.3. Метод ортогональных разложений (POD).....	334
3.3.4. Применение метода POD для сжатия данных и анализа режимов течения	338
3.3.5. Выводы по разделу	352
3.4. Обоснование применимости моделей машинного обучения. Оценка надежности и производительности моделей по сравнению с методами численного моделирования. Оценки ошибок моделей машинного обучения для решаемых задач	353
3.4.1. Модель нейронной сети для моделирования склонового потока.....	353
3.4.2. Модель нейронной сети iceMPLNet для предсказания формы льда	371
3.4.3. Описание задачи с метеостанциями и предсказанием метеопараметров	377
3.4.4. Выводы по разделу	383
3.5. Дальнейшие усовершенствования вычислительных вихревых методов. Разработка новых моделей турбулентности для вихревого метода с использованием	

методов МО. Ожидается существенное уменьшение расхождения результатов двумерного моделирования с наблюдаемыми в экспериментах параметрами обтекания цилиндрических тел большого удлинения.....	384
3.5.1. Обзор лагранжевых методов в гидродинамике	384
3.5.2. Моделирование методом крупных вихрей	385
3.5.3. Определяющие соотношения метода вихревых частиц	387
3.5.4. Определяющие соотношения метода вихревых частиц в общем виде ...	393
3.5.5. Модифицированный метод VPM.....	399
3.5.6. Моделирование обтекания кругового цилиндра	412
3.5.7. Алгоритмы машинного обучения	416
3.5.8. Выводы по разделу	417
3.6. Использование генеративно-состязательных сетей, в том числе и со сверточной нейронной сетью в качестве генератора и дискриминатора, для улучшения качества данных полученных в ходе математического моделирования экспериментов в области геофизической гидродинамики (сильные порывы ветра в атмосферном пограничном слое), удовлетворяющего закону Колмогорова-Обухова	418
3.6.1. Генеративные-состязательные сети с учетом физических законов для турбулентного течения потока	418
3.6.2. Имплементация методов коррекции ошибок и получение надежных систем ИИ с объясняемыми решениями для задач изучения опасных природных явлений	440
3.6.3. Заключение.....	455
3.7. Обработка и анализ электронной базы данных больных и реконвалесцентов новой коронавирусной инфекции	458
3.7.1. Введение.....	458
3.7.2. Материалы и методы исследования	460
3.7.3. Показатели эндотелиальной дисфункции и их взаимосвязь с гемастазнологическими/коагулологическими показателями у пациентов с COVID-19 в острый период болезни	466
3.7.4. Показатели эндотелиальной дисфункции и их взаимосвязь с клиническим течением постковидного периода у реконвалесцентов COVID-19	471

3.7.5. Реологические свойства эритроцитов у пациентов с новой коронавирусной инфекцией.....	475
3.7.6. Диагностическая значимость определения протеаз тучных клеток в прогнозировании тяжести эндотелиальной дисфункции и клинического течения SARS-CoV-2	481
3.7.7. Напряженность специфического гуморального иммунного ответа у пациентов с новой коронавирусной инфекцией в клинической форме пневмонии, и факторы, ассоциированные с ней.....	486
3.7.8. Заключение.....	513
3.8. Разработка методов и алгоритмов динамической идентификации (по данным) структуры графа суперкомпьютерной реализации композитных моделей в условиях неопределенности	515
3.8.1. Введение.....	515
3.8.2. Методы моделирования процессов реального мира в том числе эпидемиологические модели	518
3.8.3. Алгоритм динамической идентификации графа структуры модели в виде интегро-дифференциальных уравнений.....	518
3.8.4. Предсказание временных рядов с помощью алгоритма генеративного дизайна композитных моделей	524
3.9. Разработка программной библиотеки динамической идентификации (по данным) структуры графа суперкомпьютерной реализации композитных моделей в условиях неопределенности	525
3.9.1. Структура библиотеки	525
3.9.2. Кэширование реализаций композитных моделей в ходе динамической идентификации	527
3.9.3. Параллельная реализация эволюционной динамической идентификации	529
3.9.4. Удалённое обучение композитных моделей.....	532
3.10. Аprobация разработанных алгоритмов и программной библиотеки динамической идентификации структуры графа на примере задачи распространения опасных инфекций.....	535

3.10.1. Многомасштабная классификация моделей опасных инфекций.....	535
3.10.2. Алгоритм обучения многомерных условных распределений в виде байесовских сетей.....	536
3.10.3. Апробация алгоритма на примере моделирования течения инфекции на масштабе отдельно взятого госпиталя	539
3.10.4. Заключение.....	548
ЗАКЛЮЧЕНИЕ.....	550
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	564
Приложение А.....	598
Приложение Б	600
Приложение В.....	602
Приложение Г	610
Приложение Д.....	614

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящем отчете о НИР применяют следующие сокращения и обозначения:

МО – Машинное Обучение

ИИ – Искусственный Интеллект

ReLU – Rectified Linear Unit

ResNet – Residual Neural Network

RGB – Red, Green, Blue

VGG – Visual Geometry Group

VC размерность – Размерность Вапника-Червоненкиса

ROC – Receiver Operating Characteristic

DAPCA – Domain Adaptation Principal Component Analysis

PCA – Principal Component Analysis

PQSQ – Piece-wise Quadratic Sub-Quadratic

VIIRS DNB – Visible Infrared Imaging Radiometer Suite Day/Night Band

ВР – Высокое Разрешение

СР – Сверхразрешение

НР – Низкое Разрешение

НBASE – Human Built-up And Settlement Extent

КЗС – Красный, Зеленый, Синий

BN – Batch Normalization

ARCGIS – Aeronautical Reconnaissance Coverage Geographic Information System

МГИНС – Многокомпонентная Гиперсферная Интеллектуальная Нейронная Сеть

АГП – Алгоритм Глобального Поиска

OpenFOAM – Open Source Field Operation And Manipulation

kNN – k-Nearest Neighbors

kNN-KD – k-Nearest Neighbors based on K-Dimensional Trees

kNN-ME – k-Nearest Neighbors based on Peano-Hilbert curves

ДНК – Дезоксирибонуклеиновая Кислота

EWAS – Epigenome-Wide Association Study

TCGA – The Cancer Genome Atlas

GEO – Gene Expression Omnibus

IPD – Individual Participant Data

DMP – Differentially Methylated Positions

DMR – Differentially Methylated Regions